Bayesian Network Classifiers for Categorizing Cortical GABAergic Interneurons

Bojan Mihaljević · Ruth Benavides-Piccione · Concha Bielza · Javier DeFelipe · Pedro Larrañaga

Abstract An accepted classification of GABAergic interneurons of the cerebral cortex is a major goal in neuroscience. A recently proposed taxonomy based on patterns of axonal arborization promises to be a pragmatic method for achieving this goal. It involves characterizing interneurons according to five axonal arborization features, called F1–F5, and classifying them into a set of predefined types, most of which are established in the literature. Unfortunately, there is little consensus among expert neuroscientists regarding the morphological definitions of some of the proposed types. While supervised classifiers were able to categorize the interneurons in accordance with experts' assignments, their accuracy was limited because they were trained with disputed labels. Thus, here we automatically classify interneuron subsets with different label reliability thresholds (i.e., such that every cell's label is backed by at least a certain (threshold) number of experts). We quantify the cells with parameters of axonal and dendritic morphologies and, in order to predict the type, also with axonal features F1–F4 provided by the experts. Using Bayesian network classifiers, we accurately characterize and classify the interneurons and identify useful predictor variables. In particular, we discriminate among reliable examples of common basket, horse-tail, large basket, and Martinotti cells with up to 89.52 % accuracy, and single out the number of branches at 180 μm from the soma, the convex hull 2D area, and axonal features F1–F4 as especially useful predictors for distinguishing among these types. These results open up new possibilities for an objective and pragmatic classification of interneurons.

Keywords Neuronal classification · Morphological features · Label reliability · Multiple annotators · Weighted naive Bayes

Introduction

GABAergic interneurons of the cerebral cortex are highly diverse with regard to their morphological, physiological and molecular features (Ascoli et al. 2008). The scientific community considers that this variability indicates that different classes of interneurons exist and that the differences between classes are functionally relevant. Although a number of classification schemes have been proposed since the times of Santiago Ramón y Cajal (e.g., Fairén et al. 1992; Kawaguchi 1993; Cauli et al. 1997; Somogyi et al. 1998; Gupta et al. 2000; Maccaferri and Lacaille 2003), there is no catalogue of interneuron types accepted by the scientific community (DeFelipe et al. 2013), making it hard to organize data and share knowledge. The Petilla terminology

(Ascoli et al. 2008) provided a stepping stone towards a future classification by standardizing the nomenclature of features that distinguish types of GABAergic interneurons. However, the morphological, physiological, and molecular classification schemes derived from the terminology are incomplete on their own while their comprehensive combination is difficult to realize in practice (DeFelipe et al. 2013).

Thus, as a pragmatic alternative, DeFelipe et al. (2013) proposed a taxonomy based mainly on patterns of axonal arborization. It involves characterizing interneurons according to five axonal arborization features, such as the distribution of the arbor with respect to the soma's layer, as well as classifying them into a set of predefined types, most of which are established in the literature. The authors sought community-wide consensus on the taxonomy and they, therefore, asked 42 expert neuroscientists to categorize a representative set of interneurons. Unfortunately, there was little inter-expert agreement on the morphological definitions of some of the proposed types; however, there was, in general, consensus on the definitions of the axonal features. As a positive outcome, supervised classifiers were able to categorize the interneurons in accordance with the experts' assignments. Nonetheless, they were inaccurate at predicting the interneuron type and one of the axonal features; the former result being reasonable since distinguishing interneuron types was hard even for the experts. The classifiers were trained by labelling each cell with the category that was most commonly selected by the experts (majority vote). While this is a common approach in supervised classification with multiple annotators (Raykar et al. 2010) it allowed labels that were backed by few experts (i.e., 'weak majorities', those due to very ambiguous categorizations) to be used as true labels.

Thus, in this paper, we approach supervised classification of interneurons by using reliably labeled cells alone. The rationale is that unreliable labels would confuse a classifier and that it would be better to have them assigned by a model trained with reliable labels. One possibility would be to remove unreliable labels (i.e., unlabel instances) and cluster the data in a semi-supervised fashion, similarly to Mihaljević et al. (2014), allowing the discovery of previously unknown types. Here we classify reliably labeled cells alone, in a supervised fashion, seeking to discover whether at least a subset of interneurons can be accurately categorized. Rather than proposing a criterion for discerning reliable labels from unreliable ones, we form many data subsets with different label reliability thresholds (i.e., such that each instance's label is backed by at least a certain (threshold) number of experts) and analyze the supervised classification of each data subset separately.

We separately categorize the interneurons according to interneuron type and four axonal features,[1] termed F1, F2, F3, and F4 (see 'Materials and Methods' for details). We measure 214 parameters of axonal and dendritic arborizations and use all or some of them as predictive variables, depending on the axonal feature/type to be predicted. Additionally, we use axonal features F1–F4 as predictors of the interneuron type, both on their own and together with the morphological parameters. We estimate their crisp (i.e., atomic) values with majority vote, thus discarding, when using axonal features F1–F4 as predictors, interneurons unreliably categorized according to at least one of those features. Figure 1 shows an overview of the described supervised classification tasks.

We tackle each classification task with four different Bayesian network classifiers (Bielza and Larrañaga 2014). These are competitive performance classifiers (Morales et al. 2013; Friedman et al. 1997) that allow for analyzing probabilistic relationships among the variables of a domain and, among other desirable properties, accommodate for feature selection, can fit complex labelling scenarios—such as missing or partial labels; multiple class variables—and can be efficiently learned from data (see Bielza and Larrañaga 2014).

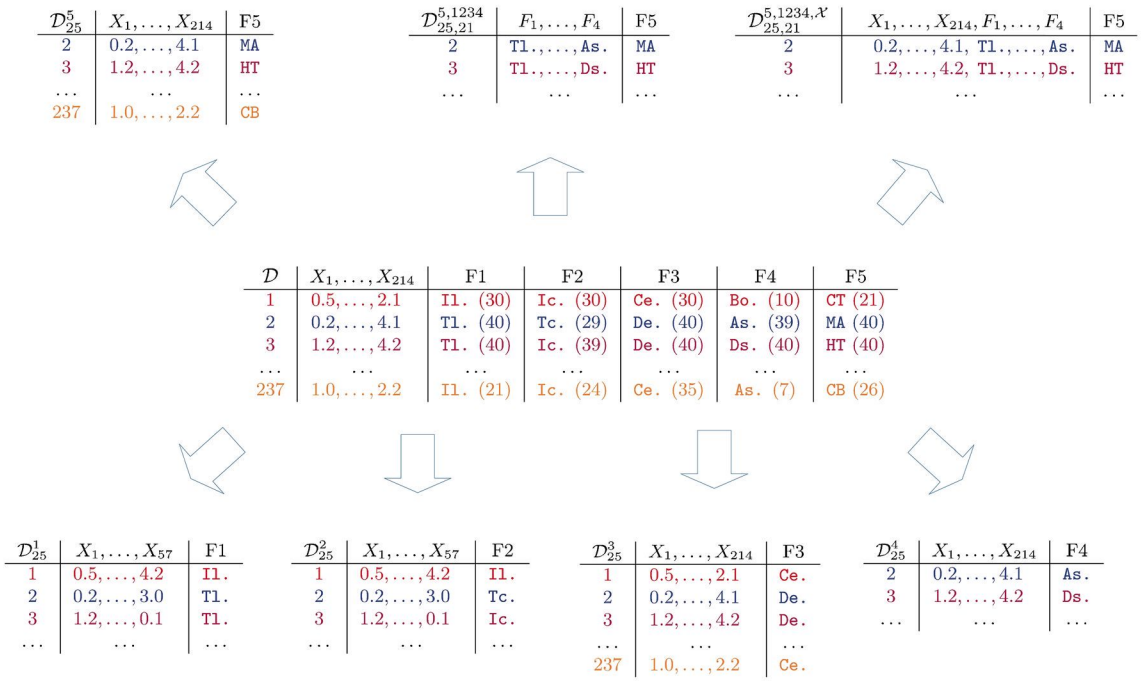
The rest of this paper is organized as follows. 'Materials and Methods' describes the data and the practical approach to interneuron classification and then elaborates on the methodology—the formation of data sets according to label reliability; the Bayesian network classifiers used; data preprocessing; and the empirical setup. 'Results' presents the results and 'Conclusions' rounds off with conclusions.

Materials and Methods

Data

All neurons and experts' terminological choices were gathered by DeFelipe et al. (2013). We used 241 cells from different areas and layers of the cerebral cortex of the mouse, rat, and monkey, whose digital reconstructions were obtained from NeuroMorpho.Org (Ascoli et al. 2007). 40 digital reconstructions exhibited an interrupted axonal process; if deemed feasible (36 cells), we drew the small missing fragments using the Neurolucida workstation (Glaser and Glaser 1990), thus completing the axon. Due to non-trivially reconstructible axons four cells were omitted from our study, yielding the final sample of 237 cells.

[1]The prediction of one of the features, F6, is almost trivial and was thus not considered here.



$\mathcal{D}^{5}_{25}$	$X_1,\ldots,X_{214}$	F5
2	0.2, ..., 4.1	MA
3	1.2, ..., 4.2	HT
...	...	...
237	1.0, ..., 2.2	CB

$\mathcal{D}^{5,1234}_{25,21}$	$F_1,\ldots,F_4$	F5
2	Tl., ..., As.	MA
3	Tl., ..., Ds.	HT
...	...	...

$\mathcal{D}^{5,1234,\mathcal{X}}_{25,21}$	$X_1,\ldots,X_{214},F_1,\ldots,F_4$	F5
2	0.2, ..., 4.1, Tl., ..., As.	MA
3	1.2, ..., 4.2, Tl., ..., Ds.	HT
...	...	...

$\mathcal{D}$	$X_1,\ldots,X_{214}$	F1	F2	F3	F4	F5
1	0.5, ..., 2.1	Il. (30)	Ic. (30)	Ce. (30)	Bo. (10)	CT (21)
2	0.2, ..., 4.1	Tl. (40)	Tc. (29)	De. (40)	As. (39)	MA (40)
3	1.2, ..., 4.2	Tl. (40)	Ic. (39)	De. (40)	Ds. (40)	HT (40)
...	...	...	...	...	...	...
237	1.0, ..., 2.2	Il. (21)	Ic. (24)	Ce. (35)	As. (7)	CB (26)

$\mathcal{D}^{1}_{25}$	$X_1,\ldots,X_{57}$	F1
1	0.5, ..., 4.2	Il.
2	0.2, ..., 3.0	Tl.
3	1.2, ..., 0.1	Tl.
...	...	...

$\mathcal{D}^{2}_{25}$	$X_1,\ldots,X_{57}$	F2
1	0.5, ..., 4.2	Il.
2	0.2, ..., 3.0	Tc.
3	1.2, ..., 0.1	Ic.
...	...	...

$\mathcal{D}^{3}_{25}$	$X_1,\ldots,X_{214}$	F3
1	0.5, ..., 2.1	Ce.
2	0.2, ..., 4.1	De.
3	1.2, ..., 4.2	De.
...	...	...
237	1.0, ..., 2.2	Ce.

$\mathcal{D}^{4}_{25}$	$X_1,\ldots,X_{214}$	F4
2	0.2, ..., 4.1	As.
3	1.2, ..., 4.2	Ds.
...	...	...

Fig. 1 A schematic overview of our automatic categorization of interneurons according to type and axonal features F1–F4. The full data set, $\mathscr{D}$ (*center*), is used to form data subsets (shown around $\mathscr{D}$) with different class variables (indicated by the superscript; e.g., F1 is the class variable in $\mathscr{D}^{1}_{25}$), predictor variables (second and third superscript; e.g., $\mathscr{D}^{5,1234}_{25,21}$ has features F1–F4 as predictors of the type), label reliability threshold (first subscript; e.g., $\mathscr{D}^{1}_{25}$ has label reliability 25), and predictor reliability threshold (second subscript; e.g., $\mathscr{D}^{5,1234}_{25,21}$ has reliability 21 for predictors F1–F4). In $\mathscr{D}$, all instances are quantified with 214 morphological parameters and labeled with the majority vote for the interneuron type and each axonal feature, with the label reliability (number of agreeing experts) shown in parentheses. The predictive variables in the classification tasks are indicated by the columns of the corresponding data sets (e.g., the predictors for F1 are the morphological parameters X_1–X_{57}; see $\mathscr{D}^{1}_{25}$). Note how label and predictor reliability determines which instances are included in a data set: for example, the first instance in $\mathscr{D}$ (shown in *red*) is omitted from $\mathscr{D}^{5}_{25}$ because its label reliability is 21, i.e., it is not above 25, the label reliability threshold in $\mathscr{D}^{5}_{25}$. Likewise, instance 237 in $\mathscr{D}$ (shown in *orange*) is omitted from $\mathscr{D}^{5,1234}_{25,21}$ and $\mathscr{D}^{5,1234,\mathcal{X}}_{25,21}$ because its reliability for F1 is not above 21. Besides the label reliability thresholds depicted here, many others were considered for each categorization task, e.g., $\mathscr{D}^{1}_{24}$, $\mathscr{D}^{1}_{28}$, $\mathscr{D}^{5,1234,\mathcal{X}}_{30,21}$, etc

Axonal Feature-Based Nomenclature

The interneuron nomenclature proposal by DeFelipe et al. (2013) consists of six categorical features of axonal arborization. These features have the following categories:

- Feature 1 (F1): `intralaminar` and `translaminar`
- Feature 2 (F2): `intracolumnar` and `transcolumnar`
- Feature 3 (F3): `centered` and `displaced`
- Feature 4 (F4): `ascending`, `descending`, and `both`
- Feature 5 (F5): `arcade` (`AR`), `Cajal-Retzius` (`CR`), `chandelier` (`CH`), `common basket` (`CB`), `common type` (`CT`), `horse-tail` (`HT`), `large basket` (`LB`), `Martinotti` (`MA`), `neurogliaform` (`NG`), and `other` (`OT`)
- Feature 6 (F6): `characterized` and `uncharacterized`

Axonal features F1 and F2 refer to the distribution of the axonal arbor relative to the cortical layer and column of the soma, respectively. Cells with the axon predominantly in the soma's cortical layer are `intralaminar`, whereas the rest are `translaminar`. Likewise, regarding F2, cells with the axon mainly confined to the soma's cortical column are `intracolumnar`; the rest are `transcolumnar` (the column is assumed to have a diameter of 300 μm). Feature F3 refers to the relative location of axonal and dendritic arbors. Cells with the dendritic arbor mainly located in the center of the axonal arborization are `centered` whereas the rest are `displaced`. Feature F4 allows for further distinguishing between `translaminar` and `displaced` cells: cells with an axon mainly ascending towards the cortical surface are `ascending`, cells with an axon mainly descending towards the white matter are `descending`, whereas the rest are termed `both`. Feature F5 is the interneuron type. Eight of the ten types can be found in the literature (Peters and Jones 1984), whereas `common type`

was introduced by DeFelipe et al. (2013), and `other` was meant to be chosen when a type missing from the proposal was considered most adequate. Regarding F6, a cell is `uncharacterized` if its reconstruction does not allow for the characterization according to the remaining features, due to, e.g., insufficient axonal reconstruction; otherwise, a cell is `characterized`. Supervised models can predict F6 with high accuracy (DeFelipe et al. 2013) so we do not consider this feature in this paper. Figure 2 shows two interneurons characterized according to axonal features F1–F5.

In DeFelipe et al. (2013), expert neuroscientists categorized interneurons by observing images such as those in Fig. 2. In addition, they were told the neuron's cortical layer—along with its approximate thickness—, the cortical area and species of the animal. A gray vertical shadow marked the 300 μm-wide cortical column, as shown in Fig. 2.

Expert Categorization Reliability

Each cell was categorized according to every feature by up to 42 experts. To crisply categorize a cell according to an axonal feature f, we reduced the vector of experts' choices for f to its mode (majority vote). We used such crisp categorizations as values of the class variable (labels) and of the predictor variables—axonal features F1–F4 were used as predictors of interneuron type. Cells with no unique majority vote for a feature f were discarded from classification tasks that involved f, either as the class or as a predictor variable (e.g., a cell without a unique majority vote for F4 was omitted when predicting F4 and when using F4 as a predictor of the type; it was used in all other classification tasks, e.g., when predicting F2 from the morphological variables ($\mathscr{D}^2_{25}$ in Fig. 1)). Furthermore, we formed data subsets with different label reliability thresholds, i.e., such that each instance's label was backed by at least a certain (threshold) number of experts. Thus, a data set $\mathscr{D}^f_t$, for predicting f, was formed of cells with label reliability larger than t for feature f, with $t \in \{0, \ldots, 41\}$. When using features F1–F4 as predictors, cells were additionally filtered according to reliability for F1–F4. Thus, a data set $\mathscr{D}^{5,1234}_{t,r}$, for predicting the type with features F1–F4 as predictors, is formed of cells with reliability greater than r for each of the features F1–F4, and reliability greater than t for the interneuron type (the label). A data set $\mathscr{D}^{5,1234,\mathscr{X}}_{t,r}$, with both the morphological parameters and axonal features F1–F4 as predictors of the type, is formed in the same way as $\mathscr{D}^{5,1234}_{t,r}$. When using F1–F4 as predictors, we augmented F4 with a category called `none`, to describe the cells which most experts considered as not categorizable according to F4—these cells would have otherwise been discarded due few experts having categorized them according to F4, yielding a low reliability for F4. Although this might lead to incorrect categorizations—a cell being `translaminar`, `displaced` and neither `ascending`, `descending` or `both` but instead `none`, in F1, F3, and F4, respectively—such combinations barely appeared in our data (see Fig. 3).

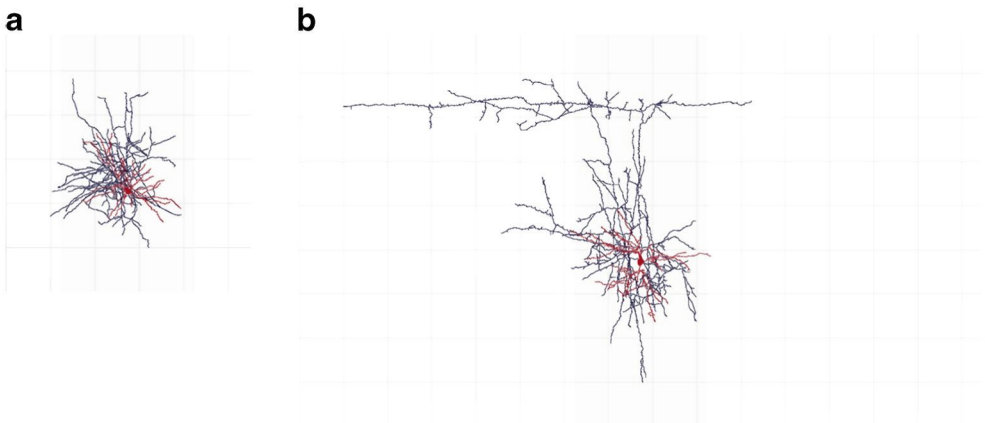


Fig. 2 Examples of interneurons of different types and with different axonal features. **a** is an `intralaminar`, `intracolumnar`, and `centered` cell, according to 38 (out of 42) experts. Most of its axon (shown in *blue*) is located within 200 μm from the soma (shown in *red*; the grid lines are established every 100 μm), thus appearing to be in the same layer as the soma; it is within the soma's cortical column (the *gray vertical shadows* depict a 300 μm-wide cortical column); and it seems to be centered around the dendritic arbor (also shown in *red*). Because this cell is not `translaminar` and `displaced`, but rather `intralaminar` and `centered`, it is not characterizable according to axonal feature F4. According to 24 experts, this is a `common basket` cell. **b** is a `translaminar`, `transcolumnar`, `displaced`, and `ascending` cell according to 38 (out of 42) experts. Unlike (**a**), this cell's axon reaches over 300 μm horizontally above soma (i.e., it seems to extend to another layer); a large portion of its axon is outside of the soma's cortical column; its dendrites are not in the center of the axonal arborization; and its axon is predominantly above the soma. According to 29 experts, this is a `Martinotti` cell

	F1	F2	F3	F4
Expert 1	`translaminar`	...	`displaced`	`descending`
Expert 2	`intralaminar`	...	`displaced`	`none`
Expert 3	`translaminar`	...	`centered`	`none`
Consensus	`translaminar`	...	`displaced`	`none`

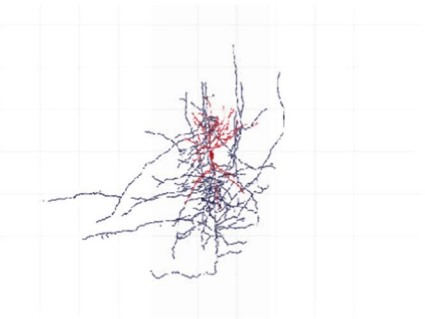


Fig. 3 An example of a theoretically invalid characterization arising due to majority vote categorization—a cell might be categorized as `translaminar` and `displaced` in F1 and F3 but neither as `ascending`, `descending`, nor `both`, but instead as `none`, in F4. The table shows a hypothetical categorization of a cell by three experts. While `translaminar` and `displaced` are the modes for F1 and F3, only one expert selected them simultaneously, and therefore only he/she also categorized the cell according to F4. The rest found that the cell was not characterisable according to F4, which was registered with the value `none`. `none` was, therefore, the mode for F4. In our data, there were only three such improperly categorized cells with reliability threshold 21 applied to axonal features F1–F4. One of them is the `CT` cell (according to 21 experts) shown on the right, characterized as `translaminar`, `intracolumnar`, `displaced`, and `none`

Morphological Variables

We used Neurolucida Explorer, the data analysis companion to Neurolucida, to compute 214 parameters of dendritic and axonal morphology using, among others, the following morphological analyses:

- Vertex analysis (described in Sadler and Berry 1983): the count of three types of bifurcations—those with two terminal branches attached (V_a); with one terminal and one bifurcating branch attached (V_b); and with two bifurcating branches attached (V_c).
- Convex hull analysis: various measures of how much space the arbor occupies.
- Sholl analysis: a histogram of intersections of the arbor and a series of concentric spheres centered at the soma. Besides the intersections, we computed histograms of the endings, nodes, and arbor length between two contiguous spheres.
- Fractal analysis (described in Panico and Sterling 1995): box-counting k-dimension—a measure of how well the arbor fills the space.
- Fan-in analysis (Glaser and McMullen 1984): torsion ratio—a measure indicative of any preferred orientation of the arbor.
- Polar histogram (McMullen et al. 1984): a round directional histogram of total arborization length corresponding to an angle interval.

Table 1 shows all the parameters that we computed, grouped by morphological analysis. We applied each analysis, except for 'Dendritic analysis', to both the axon and the dendrites. So, for example, we computed the torsion ratio of the axon and the torsion ratio of the dendrites. In total, we computed 128 axonal and 86 dendritic parameters.

We used all the computed parameters as predictive variables for predicting axonal features F3, F4, and F5. For predicting features F1 and F2 we used only the following 57 axonal parameters: total length ($\mathbf{B}_3$), number of endings ($\mathbf{B}_1$), mean branch length ($\mathbf{B}_6$), torsion ratio ($\mathbf{FI}_1$), convex hull parameters ($\mathbf{C}_1$–$\mathbf{C}_4$), Sholl analysis of intersections starting from radius 240 μm ($\mathbf{S}_4$–$\mathbf{S}_{16}$), and polar histogram ($\mathbf{P}_1$–$\mathbf{P}_{36}$).

Discrete Bayesian Network Classifiers

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the vector of n discrete predictor random variables (or attributes), let C be the (discrete) class random variable, and let $\mathscr{D} = \{(\mathbf{x}^{(1)}, c^{(1)}), \ldots, (\mathbf{x}^{(N)}, c^{(N)})\}$ be the data sample. A *Bayes classifier* estimates the probability distribution $p(\mathbf{x}, c)$ from $\mathscr{D}$ and upon classifying it assigns an instance $\mathbf{x}$ to the most probable a posteriori class,

$$c^* = \arg\max_c p(c|\mathbf{x}) = \arg\max_c p(\mathbf{x}, c).$$

Unfortunately, $p(\mathbf{x}, c)$ is seldom feasible to estimate because it has exponentially many parameters in n. Instead, it is commonly approximated by assuming conditional independences among the predictor variables. Using Bayesian networks (Pearl 1988), even distributions over many variables can be compactly encoded when sufficient conditional independence assumptions are made. Models built on strong independence assumptions, such as the *naive Bayes* (Minsky 1961), exhibit low variance and are thus especially useful for high dimension small sample data, whereas models that relax these assumptions, such as the *tree augmented naive Bayes* (Friedman et al. 1997), can outperform them when data suffices to estimate the increased number of parameters. We use examples of both of these types of Bayesian network classifiers; they are described below.

The naive Bayes (Fig. 4a) is a simple but effective classifier (Hand and Yu 2001). It approximates $p(c|\mathbf{x})$ by

Table 1 Axonal and dendritic parameters, grouped by morphological analysis

Analysis	Parameters	Axon Examples
Branching	$\mathbf{B}_1$–$\mathbf{B}_2$ Number of endings and bifurcations	$\mathbf{B}_1 = 962.00$
	$\mathbf{B}_3$–$\mathbf{B}_4$ Total and mean arbor length	$\mathbf{B}_3 = 41{,}697.10\ \mu m$
	$\mathbf{B}_5$–$\mathbf{B}_8$ Total, mean, median and std. deviation of branch length	$\mathbf{B}_6 = 21.75\ \mu m$
	$\mathbf{B}_9$ Highest branch order	$\mathbf{B}_9 = 103.00$
Convex hull	$\mathbf{C}_1$–$\mathbf{C}_2$ Area and perimeter of 2D convex hull	$\mathbf{C}_1 = 174{,}185.00\ \mu m^2$
	$\mathbf{C}_3$–$\mathbf{C}_4$ Volume and surface of 3D convex hull	$\mathbf{C}_3 = 18{,}864{,}800.00\ \mu m^3$
Sholl	Spheres of radii $\{60\ \mu m, 2 \times 60\ \mu m, \ldots, R \times 60\ \mu m\}$, with $R = 16$ for the axon and $R = 5$ for the dendrites.	
	$\mathbf{S}_1$–$\mathbf{S}_R$ Intersections with the R spheres	$\mathbf{S}_1 = 57.00$
	$\mathbf{S}_{R+1}$–$\mathbf{S}_{2R}$ Endings within the R spheres	$\mathbf{S}_{17} = 84.00$
	$\mathbf{S}_{2R+1}$–$\mathbf{S}_{3R}$ Nodes within the R spheres	$\mathbf{S}_{33} = 91.00$
	$\mathbf{S}_{3R+1}$–$\mathbf{S}_{4R}$ Arbor length within the R spheres	$\mathbf{S}_{49} = 3{,}528.00\ \mu m$
Fractal	$\mathbf{F}_1$ Box-counting k-dimension	$\mathbf{F}_1 = 1.49$
Vertex	$\mathbf{V}_1$–$\mathbf{V}_4$ V_a, V_b, V_c, and $\frac{V_a}{V_b}$	$\mathbf{V}_1 = 225.00$
Branch angle	$\mathbf{BA}_1$–$\mathbf{BA}_9$ Mean, median, and std. deviation of planar, local and spline bifurcation angles	$\mathbf{BA}_1 = 1.11$ rad
Fan-in	$\mathbf{FI}_1$ Torsion ratio	$\mathbf{FI}_1 = 1.16$
Polar	36 angle intervals of width 0.17 radians, starting with [0 rad, 0.17 rad).	
histogram	$\mathbf{P}_1$–$\mathbf{P}_{36}$ Length corresponding to the angle intervals	$\mathbf{P}_1 = 625.40\ \mu m$
Dendritic	$\mathbf{D}_1$ Number of first-order dendrites	Does not apply
	$\mathbf{D}_2$ Number of bifurcating first-order dendrites	

The parameter abbreviations (shown in bold) indicate the number of parameters corresponding to each analysis (e.g., there were nine branching parameters); the exception is Sholl analysis for which there were 20 dendritic and 64 axonal parameters, because more spheres were considered for the axon. The 'dendritic' analysis (lowermost column) applies only to the dendrites. The rightmost column shows examples of applying the analyses to a cell's axon

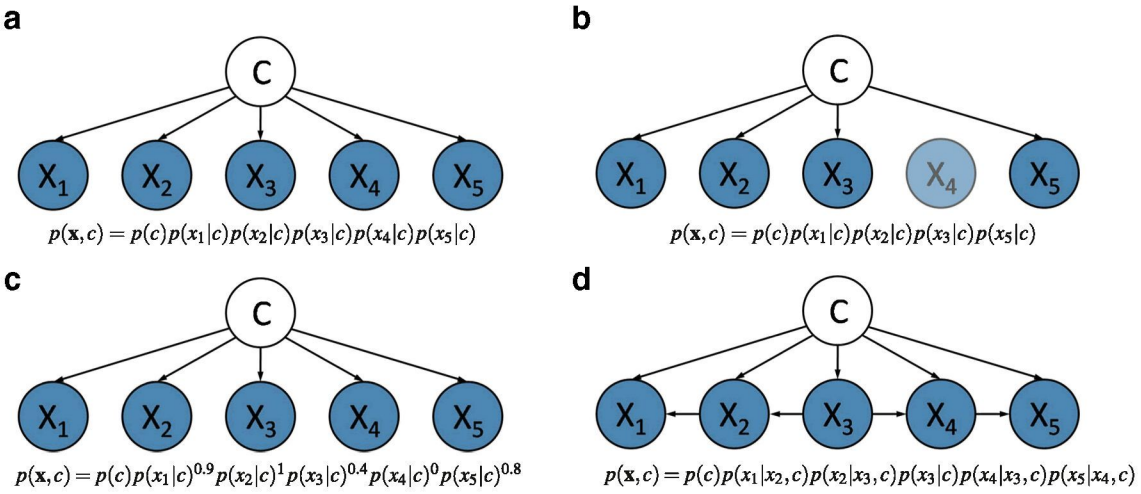


Fig. 4 Examples of the structures of the Bayesian network classifiers used in this paper. *Blue nodes* denote predictive variables whereas the *white node* denotes the class variable. The *light blue node* depicts a domain variable omitted from the model. The formulas are the factorizations of $p(\mathbf{x}, c)$ that follow from the structure above them. **a** naive Bayes: all predictors are conditionally independent given the class; **b** selective naive Bayes: predictor X_4 is omitted from the model; the remaining predictors are independent given the class; **c** attribute weighted naive Bayes: factorized like naive Bayes but, due to weights, it yields different class posterior probabilities; note that although X_4 is included in the model it is irrelevant for computing $p(\mathbf{x}, c)$ because its weight is 0; and **d** tree augmented naive Bayes, with X_3 being the root of the augmenting tree

assuming that the predictors are conditionally independent given the class, i.e.,

$$p(c|\mathbf{x}) \propto p(c) \prod_{i=1}^{n} p(x_i|c).$$

Given its fixed factorization, inducing a naive Bayes amounts to estimating the parameters of the conditional probability distributions.

Feature (predictor) selection can improve the predictive accuracy of naive Bayes (Langley and Sage 1994). Furthermore, parsimonious models can be more cost-efficient and easier to interpret. The forward sequential selection naive Bayes (NB-FSS) algorithm (Langley and Sage 1994) learns a *selective naive Bayes* (Fig. 4b) with a greedy forward wrapper (i.e., guided by accuracy) search. In other words, it starts from an empty predictor set (a model consisting solely of the class variable) and progressively incorporates predictors as long as they do not degrade predictive accuracy.

A generalization of selective naive Bayes is the *weighted naive Bayes* (Fig. 4c), given by

$$p(c|\mathbf{x}) \propto p(c) \prod_{i=1}^{n} p(x_i|c)^{w_i},$$

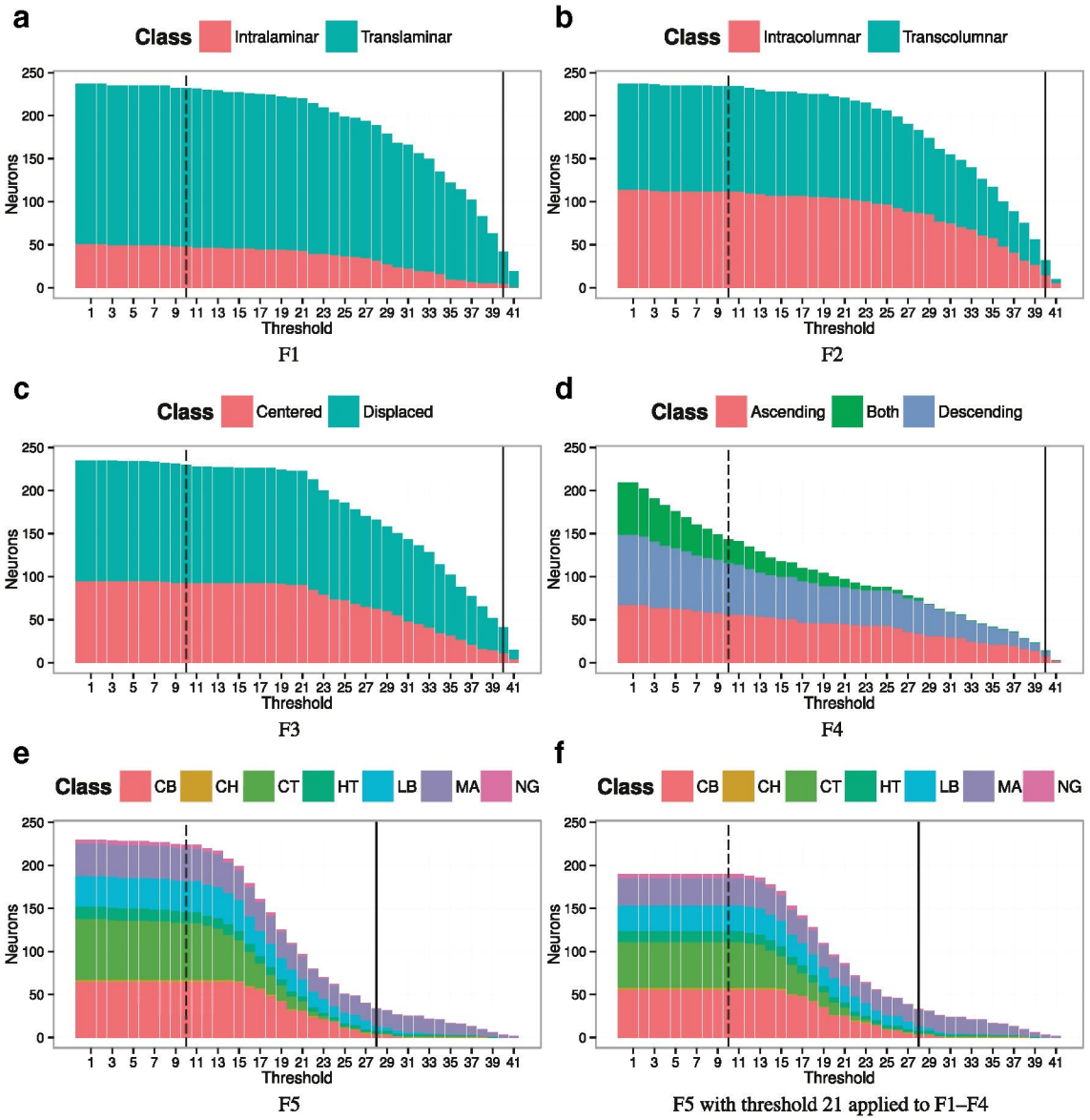


Fig. 5 Number of cells of different classes of F1–F5, versus label reliability threshold. Note that F1 is unbalanced: there are many more `translaminar` than `intralaminar` cells. Regarding F4, there are no `both` cells above threshold 28 whereas for F5 there are no `NG` cells beyond threshold 24 in both (**e**) and (**f**). *Solid vertical lines* indicate the highest label reliability threshold with no fewer than five instances of at least two classes. *Dashed vertical lines* indicate the lowest label reliability threshold considered for classification

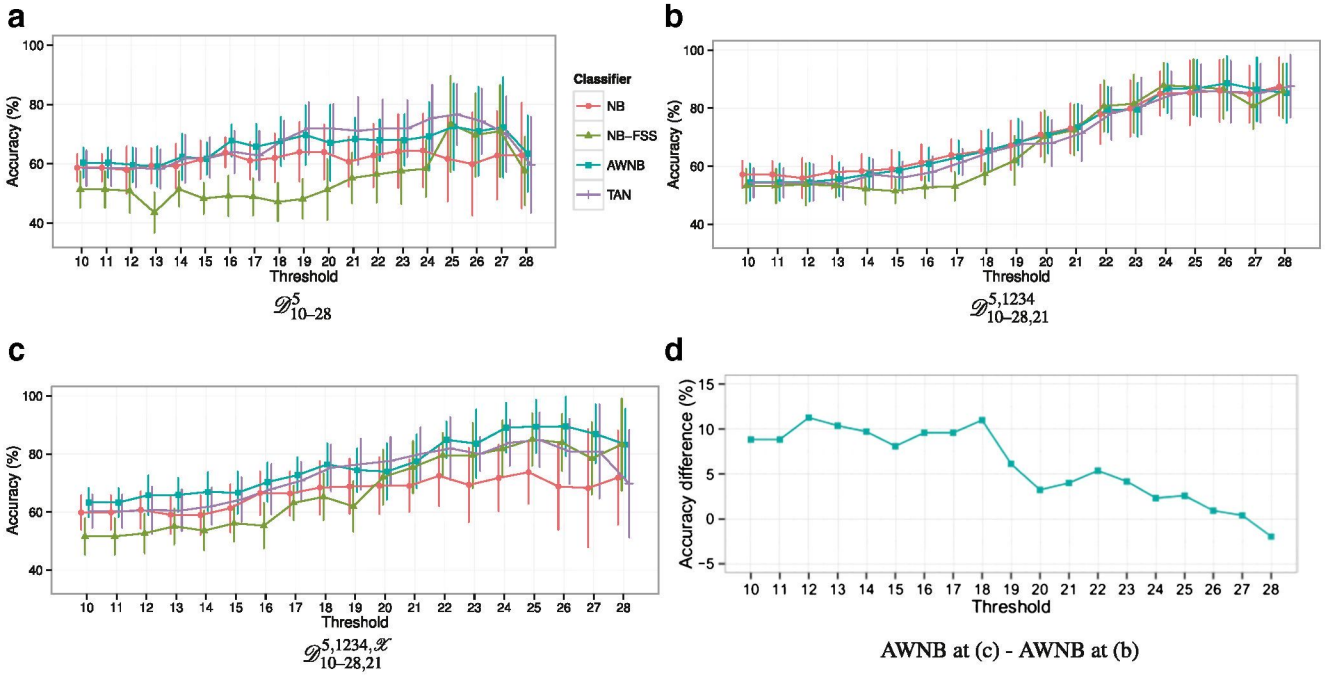


Fig. 6 Interneuron type classification accuracy versus label reliability threshold. **a** With morphological variables as predictors ($\mathscr{D}^{5}_{10\text{–}28}$); **b** with axonal features F1–F4 as predictors ($\mathscr{D}^{5,1234}_{10\text{–}28,21}$); and **c** with axonal features F1–F4 and morphological variables as predictors ($\mathscr{D}^{5,1234,\mathscr{X}}_{10\text{–}28,21}$). For (**b**) and (**c**), values of F1–F4 were obtained with a label reliability threshold 21, the used data thus being a subset of data used in (**a**). **b** and **c** were produced by identical cross-validation partitions (i.e., no differences between them due to chance). **d** plots the accuracy of AWNB at $\mathscr{D}^{5,1234,\mathscr{X}}_{10\text{–}28,21}$ minus its accuracy at $\mathscr{D}^{5,1234}_{10\text{–}28,21}$. Error bars in (**a**), (**b**), and (**c**) show the standard deviation of accuracy from five runs of five-fold cross-validation

with $w_i \in [0, 1]$ (predictor selection amounts to using weights $w_i \in \{0, 1\}$; Zaidi et al. 2013). The attribute weighted naive Bayes classifier (AWNB; Hall 2007) weights a predictor X_i in inverse proportion to X_i's dependence on other features; this dependence is estimated as $\frac{1}{\sqrt{d_i}}$, where d_i is the minimum depth at which X_i is tested in an unpruned decision tree ($\frac{1}{\sqrt{d_i}} = 0$ is assumed if X_i is not in the tree). To stabilize the estimate, various trees are constructed using bagging (Breiman 1996) and the estimates are averaged across the ensemble.

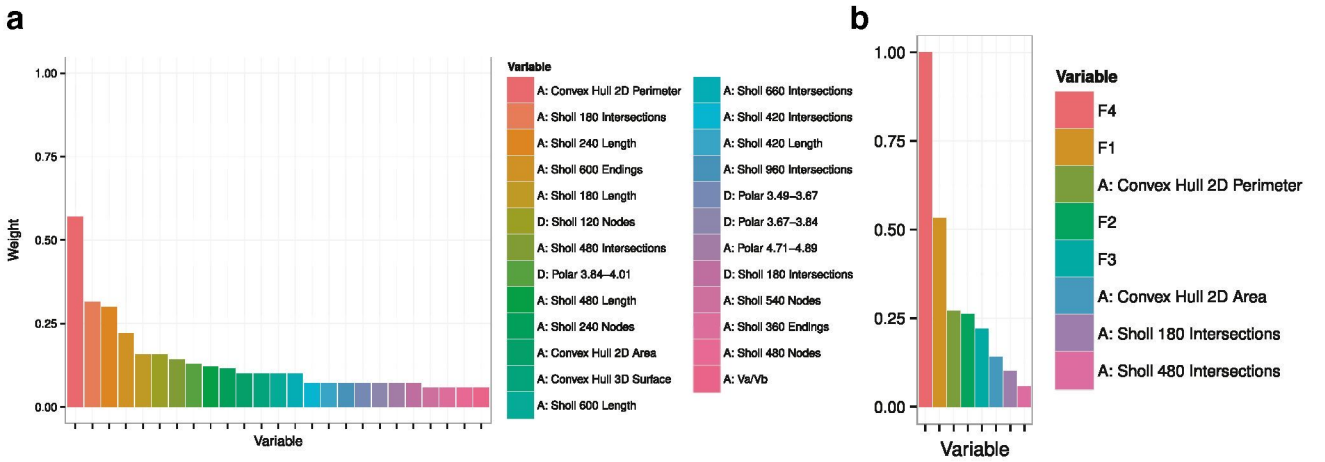


Fig. 7 AWNB weights for predictor variables at **a**: $\mathscr{D}^{5}_{25}$ and **b**: $\mathscr{D}^{5,1234,\mathscr{X}}_{26,21}$. Only variables with weights greater than zero are shown. 'A:' denotes an axonal variable whereas a 'D:' denotes a dendritic variable. The numbers in Sholl variables refer to distances of spheric rings from soma, in μm, e.g., 'A: Sholl 600 Length' is the axonal arborization length within the spheric ring at 540 μm–600 μm from soma. The numbers in the polar histogram variables refer to radian intervals, e.g., 'A: Polar 4.71–4.89' refers to axonal arborization length corresponding to the angle between 4.71 and 4.89 radians. The weights were computed by learning an AWNB model from the full data set, after selecting the 100 variables with the highest mutual information with the class variable

Table 2 Top: NB-FSS' confusion matrix for $\mathscr{D}_{25}^{5}$ (i.e., for predicting feature F5 at threshold 25; Fig. 1 explains this notation)

	CB	CH	CT	HT	LB	MA
CB	10	1			2	1
CH						
CT						
HT						
LB					9	2
MA	1		1	4	1	19
Sensitivity	0.91	0.00	0.00	0.00	0.75	0.86
Specificity	0.90	1.00	1.00	1.00	0.95	0.76

Columns denote the true classes whereas rows denote the predicted classes. Zeros were omitted. Bottom: NB-FSS' sensitivity and specificity per class. All values in the table were computed from a single run of stratified five-fold cross-validation and might not, therefore, exactly match the accuracy reported in Fig. 6a (yielded by five runs)

The tree augmented naive Bayes (Fig. 4d) alleviates some of naive Bayes' conditional independence assumptions. All predictors except for one—the root of the augmenting tree—are conditioned on a single other predictor, yielding

$$p(c|\mathbf{x}) \propto p(c)p(x_r|c)\prod_{i\neq r} p(x_i|x_{j(i)}, c),$$

where X_r is the tree root and $X_{j(i)}$ is the predictor that X_i is conditioned on. This classifier is efficiently learned by maximizing the likelihood of the resulting structure.

In summary, we used four Bayesian network classifiers of different complexities (i.e., conditional independence assumptions) and predictor selection and weighting techniques. We will refer to their learning algorithms with the following abbreviations: naive Bayes—NB; forward sequential selection naive Bayes—NB-FSS; attribute-weighted naive Bayes—AWNB; and tree augmented Bayes—TAN.

Discretization and Dimensionality Reduction

Before classifier induction, we converted all numeric variables (i.e. the morphological parameters) to categorical ones. This process, known as discretization (Yang et al. 2010), often yields more accurate naive Bayes classifiers than when assumptions such as that of normality are made about the underlying probability distributions (Dougherty et al. 1995). We used the equal-frequency discretization technique and determined the number of intervals as a function of data set size, following the weighted proportional k-interval discretization (WPKID) method (Yang and Webb 2003). The discretization process did not bias accuracy estimates as it was guided only by training data (within a cross-validation scheme): the test data were mapped, upon classification, to the intervals learned from training data.

The number of predictor variables (up to 218) was possibly too high for NB and TAN to perform well, as they do not perform predictor selection. Thus, after discretizing the training data and before inducing the classifiers (i.e., on the training set within a cross-validation scheme), we reduced the predictor set to the 100 variables with the highest mutual information with the class variable.[2] Since predictors were selected from the training set alone, this did not bias the cross-validated accuracy estimates (Smialowski et al. 2010).

Empirical Setup

Label Reliability Thresholds

The number of cells, naturally, decreased with higher label reliability. We only considered label reliability thresholds with no fewer than five instances of at least two classes, which provided the upper bounds for the reliability thresholds used: the bound was 40 for axonal features F1–F4 (see Fig. 5a–d), and 28 for F5 (Fig. 5e). The lower bound in all classification tasks was ten, roughly corresponding to one quarter of the experts.

There were seven interneuron types up to threshold 24; no NG cells remained on higher thresholds. Regardless of the threshold there were fewer than five CH, HT, and NG cells, making these types especially hard to identify. Regarding F4, no both cells remained above threshold 28. The

[2]This was not applied in classification tasks with less than 100 predictors, e.g., when predicting the interneuron type with only F1–F4 as predictor variables.

Table 3 Top: TAN's confusion matrix for $\mathscr{D}_{25}^{5}$

	CB	CH	CT	HT	LB	MA
CB	10	1			1	
CH						
CT						
HT				3		
LB					8	3
MA	1		1	1	3	19
Sensitivity	0.91	0.00	0.00	0.75	0.67	0.86
Specificity	0.95	1.00	1.00	1.00	0.92	0.79

Columns denote the true classes whereas rows denote the predicted classes. Zeros were omitted. Bottom: TAN's sensitivity and specificity per class. All values in the table were computed from a single run of stratified five-fold cross-validation and might not, therefore, exactly match the accuracy reported in Fig. 6a (yielded by five runs)

Table 4 Top: AWNB's confusion matrix for $\mathscr{D}_{26,21}^{5,1234}$

	CB	CH	CT	HT	LB	MA
CB	6		1		1	
CH						
CT						
HT		1		4		
LB	2				9	
MA					1	21
Sensitivity	0.75	0.00	0.00	1.00	0.82	1.00
Specificity	0.95	1.00	1.00	0.98	0.94	0.96

Columns denote the true classes whereas rows denote the predicted classes. Zeros were omitted. Bottom: AWNB's sensitivity and specificity per class. All values in the table were computed from a single run of stratified five-fold cross-validation and might not, therefore, exactly match the accuracy reported in Fig. 6b (obtained by five runs)

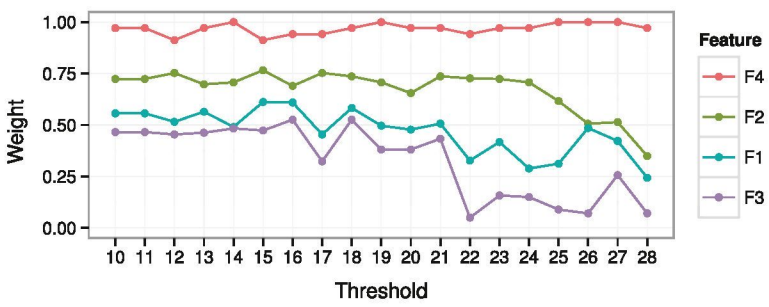


Fig. 8 AWNB predictor weights for predicting F5 from axonal features F1–F4 versus label reliability threshold. Note that the order of importance of the features is constant across the thresholds: most weight is given to F4, least to F3, with F2 and F1 in between

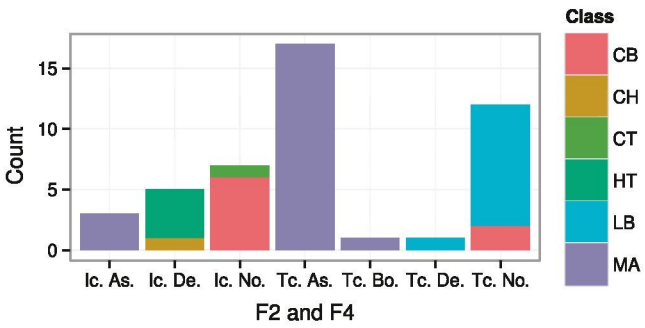


Fig. 9 Interneuron type (colours) versus combination of F2 and F4 values at threshold 26 for interneuron type and threshold 21 for F1–F4 ($\mathscr{D}_{26,21}^{5,1234}$). F2 and F4 combined could discriminate between CB, HT, LB, and MA and cells rather precisely (e.g., all transcolumnar and ascending cells were MA). Feature F4 alone separated MA from HT: all HT were descending whereas MA were either ascending or both, whereas F2 largely separated CB from LB cells: all LB were transcolumnar while most CB were intracolumnar. Abbreviations: Ic. = intracolumnar; Tc. = transcolumnar; As. = ascending; De. = descending; Bo. = both; and No. = none

predictions of F1, F2, and F3 were binary tasks regardless of label reliability threshold, i.e., the same classes were present at all thresholds considered.

When using F1–F4 as predictors of the interneuron type (F5) we fixed the reliability threshold for F1–F4 to 21—corresponding to 50 % of the experts—while considering all thresholds from 10 to 28 for F5 (see Fig. 5f), following the above-described criteria.

Classifier Parametrization and Accuracy Estimation

For NB-FSS we used resubstitution accuracy as the objective function; we halted its search process when new accuracy improved current accuracy by no more than 10 % (i.e., $acc_{\text{new}} - acc_{\text{current}} \leq 10, acc \in [0, 100]$). For the AWNB classifier, we built classification trees from 10 bootstrap samples half the size of the data set ($\frac{N}{2}$). We estimated the parameters of the Bayesian networks with Laplace correction for maximum likelihood. We estimated predictive accuracy of the classifiers with five repetitions of five-fold stratified cross-validation.

Software

The Bayesian network classifiers used are implemented in the `bnclassify` (Mihaljevic et al. 2013) package for the `R` (R Core Team 2012) statistical software environment. We used `Weka` (Hall et al. 2009) for discretization (through the `RWeka` (Hornik et al. 2009) interface for `R`) and the `caret` `R` package (Kuhn et al. 2013) for cross-validation estimation of accuracy.

Results

Predicting Interneuron Type

From Morphological Variables

In general, accuracy improved with label reliability (see Fig. 6a). Best accuracy—76.63 %—was achieved by TAN at label reliability threshold 25 ($\mathscr{D}_{25}^{5}$). While a TAN model incorporates all predictive variables,[3] AWNB and NB-FSS achieved comparable accuracy, at this threshold, with few variables: AWNB with 24 (19 axonal and five dendritic) and NB-FSS with two axonal variables: the 2D convex hull perimeter ($\mathbf{C}_2$) and Sholl intersections at 180 μm from soma ($\mathbf{S}_3$); these were also the most relevant variables according to AWNB—see Fig. 7a. At thresholds 25–27 NB-FSS was very accurate by using these two variables alone, indicating that they suffice for discriminating among CB, LB, and MA cells, which are the interneuron types that NB-FSS was able identify at these thresholds (see Table 2). Unlike NB-FSS, TAN also managed to identify HT cells at threshold 25, thus accurately discriminating among CB, HT, LB, and MA cells (see Table 3).

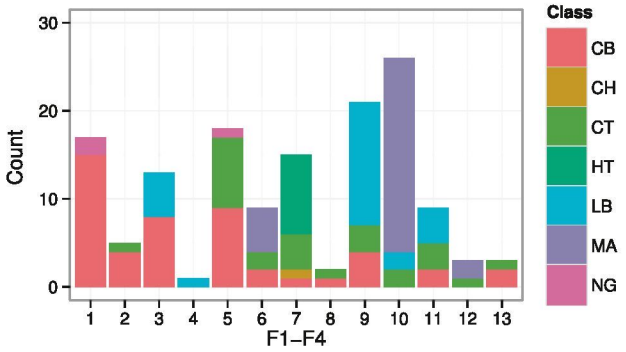
From Axonal Features F1–F4

In general, accuracy improved with label reliability (see Fig. 6b) and NB, AWNB, and TAN were similarly accurate at all thresholds. The best accuracy—88.58 %—was achieved by AWNB at threshold 26 ($\mathscr{D}_{26,21}^{5,1234}$). All classifiers could accurately discriminate between reliable examples of the CB, HT, LB, and MA types (see Table 4 for AWNB). Not only were they similarly accurate but they actually classified in a similar way—for example, TAN and NB-FSS had identical confusion matrices at threshold 26. Prediction was more accurate than with morphological predictors alone (note that, although different, the data sets from the two settings were actually similar at high reliability thresholds, see Fig. 5e and f).

F4 seemed to be the most useful axonal feature for predicting the interneuron type. Regardless of label reliability, AWNB always assigned most importance to F4, then to F2, and least to F1 and F3 (see Fig. 8). Accordingly, the NB-FSS classifier selected F4, while omitting F1 and F3, at all thresholds; when it selected F2 along with F4—at thresholds 20–27—it was more or similarly accurate as the remaining classifiers. Indeed, features F4 and F2 alone could separate reliable examples of the CB, HT, LB, and MA types (see Fig. 9). The omission of F1 and F3, in favour of F4, by NB-FSS, and their lower importance in AWNB, is reasonable since F4 by definition carries

[3]The 100 variables that were selected previous to classifier induction.

a



b

	CB	CH	CT	HT	LB	MA	NG
CB	39		11		5		3
CH							
CT							
HT	1	1	4	9			
LB	6		6		19		
MA	2		5		2	29	
NG							

Accuracy: 67.60%

c

	CB	CH	CT	HT	LB	MA	NG
CB	36		9		6		3
CH							
CT	1		1		3		
HT	1	1	4	9			
LB	7		6		15		
MA	3		6		2	29	
NG							

Accuracy: 63.38%

d

	CB	CH	CT	HT	LB	MA	NG
CB	3		2		-1		
CH							
CT	-1		-1		-3		
HT							
LB	-1				4		
MA	-1		-1				
NG							

Fig. 10 A combination of features F1–F4 does not clearly identify the interneuron type at a low reliability threshold. **a** plots the interneuron type against the combinations of features F1–F4 at $\mathscr{D}_{17,21}^{5,1234}$; **b** is the confusion matrix of an 'ideal' classifier which would assign every combination of F1–F4 values in (**a**) to its most common class. Thus, for example, the five LB cells corresponding to combination 3 (blue part of the third bar in (**a**)) would be assigned to CB (salmon-colored part of the same bar), since CB cells are predominant in combination 3. The confusion matrix of AWNB is shown in (**c**) whereas the difference between the 'ideal' confusion matrix and that of AWNB, (**b**)–(**c**), is shown in (**d**). AWNB was only slightly worse that the 'ideal' classifier: it misclassified three CB and four LB cells more than 'ideal' classifier (shown in red in (**d**)) but correctly classified one CT cell more (shown in *blue*). The latter is possible due to random permutations in cross-validation. The columns in confusion matrices denote the true classes whereas rows denote the predicted classes. Zeros were omitted. The confusion matrix of the AWNB was obtained from a single run of cross-validation. F1–F4 combinations: 1 = (intralaminar, intracolumnar, centered, none); 2 = (intralaminar, intracolumnar, displaced, none); 3 = (intralaminar, transcolumnar, centered, none); 4 = (intralaminar, transcolumnar, displaced, none); 5 = (translaminar, intracolumnar, centered, none); 6 = (translaminar, intracolumnar, displaced, ascending); 7 = (translaminar, intracolumnar, displaced, descending); 8 = (translaminar, transcolumnar, displaced, both); 9 = (translaminar, transcolumnar, centered, none); 10 = (translaminar, transcolumnar, displaced, ascending); 11 = (translaminar, transcolumnar, displaced, descending); 12 = (translaminar, transcolumnar, displaced, both); and 13 = (translaminar, transcolumnar, displaced, none)

information about F1 and F3—a cell that is `none` in F4 is not, by definition, `translaminar` and `displaced` (in F1 and F3, respectively), whereas a cell with a different F4 value (`ascending`, `descending`, or `both`) is `translaminar` and `displaced` (in F1 and F3, respectively). This redundancy of F4 on the one hand and F1 and F3 on the other might suggest that predictor weighting is an adequate approach.

Nevertheless, not even the combination of the four high-level axonal features is expressive enough to separate the types well at low reliability thresholds. *Many cells had identical axonal features F1–F4 but nevertheless belonged to different types*; since a single combination of features can only be assigned to one type, many cells cannot be correctly classified. Poor accuracies at low thresholds are partially due to this limited expressiveness; in fact, they are close to the accuracies achievable by assigning each instance to its majority class (see Fig. 10). This suggests that a richer predictor space (i.e., beyond F1–F4) might be necessary to better discriminate among interneuron types at low thresholds. For this purpose, we augmented the predictor set with the 214 morphological variables; the obtained results are presented in the following section.

From Morphological Variables and Axonal Features F1–F4

Using the morphological variables together with the high-level axonal features improved AWNB's accuracy at all thresholds except for 28 (see Fig. 6c and d). AWNB achieved the highest overall accuracy (considering all

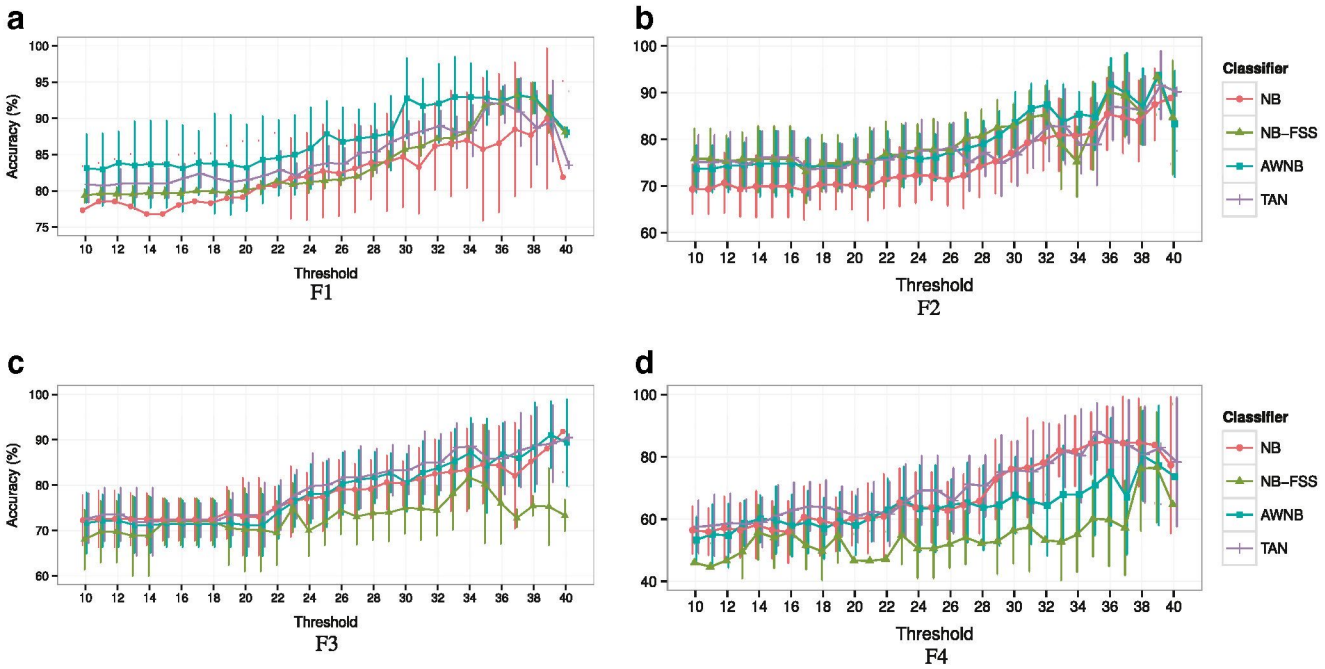


Fig. 11 Classification accuracy for axonal features F1–F4 ((**a**)–(**d**), respectively) versus label reliability threshold. Error bars show standard deviation of accuracy from five runs of five-fold cross-validation

three predictor sets) for predicting the interneuron type—89.52 %—at threshold 26 ($\mathscr{D}_{26,21}^{5,1234,\mathscr{X}}$). At this threshold, it assigned non-zero weights to eight predictors: the features F1–F4 and four morphological parameters of the axon (see Fig. 7b), with most weight assigned to F4. These morphological parameters were also used by AWNB in 'From Morphological Variables' (see Fig. 7a). In both settings most weight was assigned to the 2D convex hull perimeter ($\mathbf{C}_2$).

Augmenting the predictor set with morphological variables also improved the accuracies of TAN and NB—which do not prune the predictor set—at low thresholds (up to thresholds 22 and 19 for TAN and NB, respectively; see Fig. 6c). This seems to confirm that poor accuracies at low thresholds in the previous setting were partially due to the limited expressiveness of features F1–F4. NB and TAN performed worse at high thresholds—where features F2 and F4 suffice to discriminate among the types—, possibly because axonal features F1–F4 were dominated by the many morphological variables, only some of which seem to be useful for class prediction.

Predicting Axonal Features F1–F4

The highest accuracy for predicting the axonal feature F1—93.19 %—was achieved by the AWNB and NB-FSS classifiers at threshold 37 ($\mathscr{D}_{37}^{1}$; see Fig. 11a). These classifiers, however, assigned all cells to the predominant `translaminar` class (at threshold 37 there were 95 `translaminar` and 7 `intralaminar` cells). NB was best at correctly identifying `intralaminar` cells but had a lower accuracy that the remaining classifiers (88.43 %; see Table 5).

Regarding the prediction of the axonal feature F2, the highest accuracy—93.75 %—was achieved by AWNB at label reliability threshold 39 ($\mathscr{D}_{39}^{2}$; see Fig. 11b). It was similarly good at identifying both categories of interneurons (0.95 sensitivity and 0.91 specificity, with `intracolumnar` being the positive class). Generally, AWNB was most accurate at classifying reliably labeled cells (it was the best at thresholds 17 and 31–39) whereas NB was most often the least accurate, indicating that the predictor variables were redundant to some degree. AWNB used ten variables at threshold 39 (see Fig. 12a), with most weight assigned to convex hull 2D area ($\mathbf{C}_1$), which was the only variable selected by NB-FSS—similarly accurate

Table 5 Sensitivity, specificity and accuracy of the different classifiers at $\mathscr{D}_{37}^{1}$. `intralaminar` is considered as the positive class

	Sensitivity	Specificity	Accuracy
NB	1	0.88	88.43 %
NB-FSS	0	1	93.19 %
AWNB	0	1	93.19 %
TAN	0.11	0.97	90.78 %

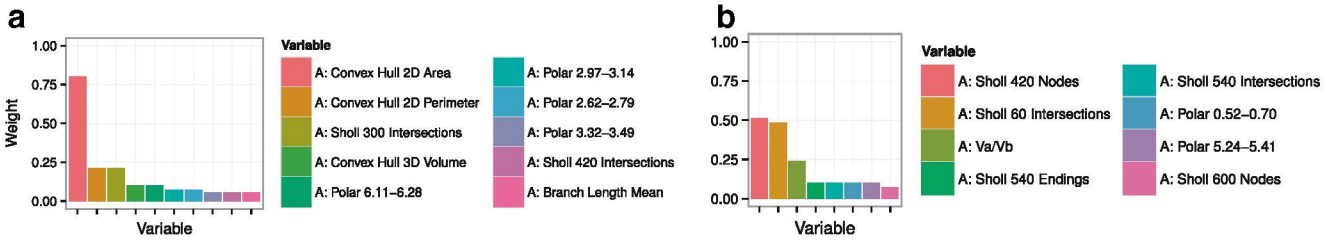


Fig. 12 AWNB weights for predictor variables at **a**: $\mathscr{D}^2_{39}$ and **b**: $\mathscr{D}^3_{40}$. Only variables with weights greater than zero are shown. 'A:' denotes an axonal variable whereas a 'D:' denotes a dendritic variable. The numbers in Sholl variables refer to distances of spheric rings from soma, in μm, e.g., 'A: Sholl 600 Length' is the axonal arborization length within the spheric ring at 540 μm–600 μm from soma. The numbers in the polar histogram variables refer to radian intervals, e.g., 'A: Polar 4.71–4.89' refers to axonal arborization length corresponding to the angle between 4.71 and 4.89 radians. The weights were computed by learning an AWNB model from the full data set, after selecting the 100 variables with the highest mutual information with the class variable

as AWNB—at this threshold. Indeed, NB-FSS selected a single convex hull variable at each threshold—either $\mathbf{C}_1$, $\mathbf{C}_3$, or $\mathbf{C}_4$—and it was nonetheless relatively accurate (scoring within 2 % of the highest accuracy at all but three thresholds).

Regarding the prediction of F3, the highest accuracy—91.83 %—was achieved by NB at label reliability threshold 40 ($\mathscr{D}^3_{40}$; see Fig. 11c). NB was slightly more accurate, at this threshold, at identifying the more numerous `displaced` cells (0.87 sensitivity and 0.93 specificity, with `centered` being the positive class). Although F4 refers to the relative distribution of axonal and dendritic arbors, AWNB was similarly accurate as NB at this threshold by using only eight axonal variables (see Fig. 12b). Although AWNB selected no more than 48 variables at a single threshold (with the number of variables inversely proportional to label reliability), it achieved similar accuracy as NB at most thresholds, which might suggest that only a subset of variables is useful for predicting the axonal feature F3.

The highest accuracy for predicting the axonal feature F4—88.10 %—was achieved by TAN at threshold 35 ($\mathscr{D}^4_{35}$; see Fig. 11d). At this threshold there were no `both` cells and the classifiers thus had to distinguish between `ascending` and `descending` cells alone. TAN was equally good at identifying both classes (0.88 sensitivity and specificity). When distinguishing between `ascending` and `descending` cells, NB and TAN outperformed the classifiers that prune the predictor set—AWNB and NB-FSS—(see Fig. 11d), which may suggest that many of the morphological variables used are useful for distinguishing among these two categories.

Conclusions

We used supervised classifiers based on Bayesian networks to automatically categorize cortical GABAergic interneurons according to their type—as commonly used in the literature—and four features of axonal arborization, called F1, F2, F3, and F4. We trained the classifiers with the categorization of 237 interneurons according to the mentioned features and the interneuron type provided by 42 expert neuroscientists. We used up to 214 morphological parameters as predictor variables. Also, we used the axonal features F1–F4 as predictors of interneuron type. Due to little inter-expert agreement on the categorization of some cells, we separately analyzed data subsets with different expert categorization reliability thresholds.

We found that the interneurons that were categorized with more inter-expert agreement were more accurately classified by our models. The models accurately distinguished between reliable examples of the common basket, horse-tail, large basket and Martinotti interneuron types. Analyzing the Bayesian network classifiers, we identified two axonal variables—the convex hull 2D perimeter and number of branches at 180 μm from soma—and the axonal features F1–F4 as especially useful for discriminating among these interneuron types. Indeed, axonal features F2 and F4 alone were able to accurately separate reliable examples of these types. Besides the interneuron type, we were also able to accurately categorize interneurons according to the axonal features F1–F4.

Since we show that axonal features F1–F4 can be accurately predicted, it might be possible to avoid recurring to experts for future neuron labellings according to these features and instead use the values provided by the models. For this purpose, it would be useful to predict the axonal features simultaneously,—using, e.g., multi-dimensional Bayesian network classifiers (Bielza et al. 2011)—instead of separately, as in the present study. Furthermore, using the predicted values of F1–F4 as predictors of the interneuron type would be an interesting problem to tackle with stacked classification or classifier chains.

As a more flexible and less subjective way to characterize the interneurons, the categorical axonal features F1–F4 might be replaced with real-valued measures: we could, for example, measure the percentage of the axon that leaves the soma's layer instead of categorically distinguishing between 'intralaminar' and 'translaminar' axons. We are therefore developing an algorithm for such objective measurement of neuronal features F1–F4.

Finally, our data consisted of neurons from several species. While it is considered that the same morphological types of neurons exist in these species, there still might be inter-species variability. In the near future, we plan to further analyze the data to take this possibility into account. Also, although our data contained seven interneuron types, three of them—chandelier, neurogliaform, and horse-tail—were represented with fewer than five cells. Since there is medium to high inter-expert consensus on the definitions of those types, it would be interesting to repeat the present analysis with more cells of these types.

Information Sharing Statement

All used data—the 237 interneuron cell reconstructions and the corresponding experts' characterizations according to features F1 to F6—are available at http://cig.fi.upm.es/bojan/gardener/. The `bnclassify` `R` package will be made available on the CRAN repository (http://cran.r-project.org/) by the end of 2014 whereas the remaining software used is publicly available: `caret` and `RWeka` on CRAN and `Weka` at http://www.cs.waikato.ac.nz/ml/weka/.

Acknowledgments This work was supported by grants from the following entities: the Spanish Ministry of Economy and Competitiveness (grants TIN2013-41592-P to B.M., C.B., and P.L.; BFU2012-34963 to J.DF.), CIBERNED CB06/05/0066 to J.DF., the Cajal Blue Brain Project (C080020-09; the Spanish partner of the Blue Brain Project initiative from EPFL) to B.M., C.B., J.DF., and P.L., and the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 604102 (Human Brain Project) to C.B., J.DF., and P.L. R.B.-P. was supported by the Spanish Ministry of Economy and Competitiveness (CSIC).

References

Ascoli, G.A., Donohue, D.E., Halavi, M. (2007). Neuromorpho.org: a central resource for neuronal morphologies. *The Journal of Neuroscience*, *27*(35), 9247–9251.

Ascoli, G.A., Alonso-Nanclares, L., Anderson, S., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., Buzsaki, G., Cauli, B., DeFelipe, J., Fairén, A., et al. (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Reviews Neuroscience*, *9*(7), 557–568.

Bielza, C., & Larrañaga, P. (2014). Discrete Bayesian network classifiers: a survey. *ACM Computing Surveys*, *47*(1), 5:1–5:43.

Bielza, C., Li, G., Larranaga, P. (2011). Multi-dimensional classification with Bayesian networks. *International Journal of Approximate Reasoning*, *52*(6), 705–727.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Cauli, B., Audinat, E., Lambolez, B., Angulo, M.C., Ropert, N., Tsuzuki, K., Hestrin, S., Rossier, J. (1997). Molecular and physiological diversity of cortical nonpyramidal cells. *The Journal of Neuroscience*, *17*(10), 3894–3906.

DeFelipe, J., López-Cruz, P.L., Benavides-Piccione, R., Bielza, C., Larrañaga, P., Anderson, S., Burkhalter, A., Cauli, B., Fairén, A., Feldmeyer, D., et al. (2013). New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nature Reviews Neuroscience*, *14*, 202–216.

Dougherty, J., Kohavi, R., Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. In *Machine learning: proceedings of the twelfth international conference* (pp. 194–202).

Fairén, A., Regidor, J., Kruger, L. (1992). The cerebral cortex of the mouse (a first contribution—the 'acoustic' cortex). *Somatosensory & Motor Research*, *9*(1), 3–36.

Friedman, N., Geiger, D., Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, *29*, 131–163.

Glaser, J.R., & Glaser, E.M. (1990). Neuron imaging with Neurolucida—a PC-based system for image combining microscopy. *Computerized Medical Imaging and Graphics*, *14*(5), 307–317.

Glaser, E.M., & McMullen, N.T. (1984). The fan-in projection method for analyzing dendrite and axon systems. *Journal of Neuroscience Methods*, *12*(1), 37–42.

Gupta, A., Wang, Y., Markram, H. (2000). Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science*, *287*(5451), 273–278.

Hall, M. (2007). A decision tree-based attribute weighting filter for naive Bayes. *Knowledge-Based Systems*, *20*(2), 120–126.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H. (2009). The WEKA data mining software: an update. *SIGKDD Explorations Newsletter*, *11*(1), 10–18.

Hand, D.J., & Yu, K. (2001). Idiot's Bayes—not so stupid after all? *International Statistical Review*, *69*(3), 385–398.

Hornik, K., Buchta, C., Zeileis, A. (2009). Open-source machine learning: `R` meets `Weka`. *Computational Statistics*, *24*(2), 225–232.

Kawaguchi, Y. (1993). Physiological, morphological, and histochemical characterization of three classes of interneurons in rat neostriatum. *The Journal of Neuroscience*, *13*(11), 4908–4923.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T. (2013). `caret`: classification and regression training. `R` package version 5.17-7.

Langley, P., & Sage, S. (1994). Induction of selective Bayesian classifiers. In *Proceedings of the 10th conference on uncertainty in artificial intelligence* (pp. 399–406): Morgan Kaufmann.

Maccaferri, G., & Lacaille, J.C. (2003). Interneuron diversity series: hippocampal interneuron classifications–making things as simple as possible, not simpler. *Trends in Neurosciences*, *26*(10), 564–571.

McMullen, N.T., Glaser, E.M., Tagamets, M. (1984). Morphometry of spine-free nonpyramidal neurons in rabbit auditory cortex. *Journal of Comparative Neurology*, *222*(3), 383–395.

Mihaljevic, B., Bielza, C., Larrañaga, P. (2013). `bayesslass`: an `R` package for learning Bayesian network classifiers. In *Proceedings of useR!—the `R` user conference* (p. 53).

Mihaljević, B., Benavides-Piccione, R., Guerra, L., DeFelipe, J., Larrañaga, P., Bielza, C. (2014). Classifying GABAergic interneurons with semi-supervised projected model-based clustering. *Artificial Intelligence in Medicine*, (in press).

Minsky, M. (1961). Steps toward artificial intelligence. *Transactions on Institute of Radio Engineers*, *49*, 8–30.

Morales, D., Vives-Gilabert, Y., Gómez-Ansón, B., Bengoetxea, E., Larrañaga, P., Bielza, C., Pagonabarraga, J., Kulisevsky, J., Corcuera-Solano, I., Delfino, M. (2013). Predicting dementia development in Parkinson's disease using Bayesian network classifiers. *Psychiatry Research: NeuroImaging*, *213*, 92–98.

Panico, J., & Sterling, P. (1995). Retinal neurons and vessels are not fractal but space-filling. *Journal of Comparative Neurology*, *361*(3), 479–490.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Francisco: Morgan Kaufmann.

Peters, A., & Jones, E.G. (1984). *Cerebral cortex: volume 1: cellular components of the cerebral cortex*. New York: Plenum Press.

R Core Team. (2012). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raykar, V.C., Yu, S., Zhao, L.H., Valadez, G.H., Florin, C., Bogoni, L., Moy, L. (2010). Learning from crowds. *The Journal of Machine Learning Research*, *11*, 1297–1322.

Sadler, M., & Berry, M. (1983). Morphometric study of the development of Purkinje cell dendritic trees in the mouse using vertex analysis. *Journal of Microscopy*, *131*(3), 341–354.

Smialowski, P., Frishman, D., Kramer, S. (2010). Pitfalls of supervised feature selection. *Bioinformatics*, *26*(3), 440–443.

Somogyi, P., Tamás, G., Lujan, R., Buhl, E.H. (1998). Salient features of synaptic organisation in the cerebral cortex. *Brain Research Reviews*, *26*(2), 113–135.

Yang, Y., & Webb, G.I. (2003). Weighted proportional k-interval discretization for naive-Bayes classifiers. In *Advances in knowledge discovery and data mining* (pp. 501–512). Springer.

Yang, Y., Webb, G.I., Wu, X. (2010). Discretization methods. In *Data mining and knowledge discovery handbook* (pp. 101–116). Springer.

Zaidi, N.A., Cerquides, J., Carman, M.J., Webb, G.I. (2013). Alleviating naive Bayes attribute independence assumption by attribute weighting. *Journal of Machine Learning Research*, *14*, 1947–1988.